SPEECH TRANSFER OVER PACKET NETWORKS USING VERY LOW DIGITAL DATA BANDWIDTHS

REFERENCE TO RELATED APPLICATIONS

[0001] Not Applicable

FIELD OF THE INVENTION

[0002] The invention relates to the communication of speech over digital networks using very low data bandwidth. Specifically, the inventive method translates speech into text at a source terminal, communicates the text across the communication link to a destination terminal, and translates the text into reproduced speech at the destination terminal.

BACKGROUND OF THE INVENTION

[0003] Telephony communications are increasingly being communicated across digital networks. As a result, it is becoming increasingly desirable to communicate voice over these networks too. Presently, this is accomplished by using Voice over Packet (VoP) systems that compress the voice in accordance with an International Telecommunication Union (ITU) standard. After being conveyed across the digital network, the voice packets are decompressed and used to reproduce a signal at the destination terminal which attempts to closely match the original signal. This solution significantly reduces the required bandwidth while maintaining high voice quality. For example, an uncompressed voice signal requires a bandwidth of 64,000 bits per second (bps). A compressed version of this same voice signal may be communicated with as little as 8,000 bps while still preserving the toll quality of the telephone call.

[0004] Each compression of an audible signal, which reduces the necessary bandwidth, also reduces the resolution of the signal and causes some distortion. A substantial bandwidth reduction below 8,000 bps may not be possible using prior art methods, without greatly affecting the quality of the voice signal.

SUMMARY OF THE INVENTION

[0005] The inventive method uses voice recognition software to translate spoken words into digitally coded text. The text is transmitted over the digital network instead of the voice signal. When the text arrives at the destination terminal, it is converted back into speech by voice recognition software. The present invention provides a means to communicate a voice signal with as little as 140 bps.

[0006] In one embodiment, a default voice is used by the destination terminal to reproduce the spoken words. However, the destination terminal may have a database of the speaker's voice profile so that the reproduced voice sounds like the voice of the original speaker. If this database does not exist at the destination terminal, it is gradually transmitted to the destination terminal during the course of the telephone conversation. Once the speaker's voice profile is completely communicated, the voice profile used by the destination terminal transitions from the default profile to the speaker's profile.

[0007] A voice profile for a speaker is can be initially generated during training of the system or can be generated during one or more conversations. Training allows for voice recognition and allows the system to build a speaker voice profile which not only enables recognition but provides voice for more natural playback at the destination. The voice profile can be created by having the speaker read text provided by the voice recognition software prior to a telephone call or through recognition and correction of speech of the user during a conversation. The software may automatically create the voice profile during a telephone conversation and store and/or transmit it as it is generated or after completion. The voice profile for a user does not have to remain static after its initial creation, it can be improved and refined as the speaker uses the system and adds to the stored vocabulary and speech. At some level of use, the system can accommodate several alternative pronunciations of words to reflect different intonations or inflections of the speaker which can then be reproduced at the destination to enhance the conversation quality. The recognition software can select between the various intonations of the word to be reproduced

based upon the recognition of the current pronunciation of the words in the speaker's present conversation.

[0008] The present invention teaches a method of communicating speech across a communication link using very low digital data bandwidth, having the steps of: translating speech into text at a source terminal; communicating the text across the communication link to a destination terminal; and translating the text into reproduced speech at the destination terminal. The primary advantage of this invention is the significant reduction of bandwidth it requires to communicate speech across a digital network and the elimination of distortion which can result from compression or packet loss.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] Preferred embodiments of the invention are discussed hereinafter in reference to the drawings, in which:

[0010] Figure 1 - illustrates a process flow executed by the source terminal of either one of two idealized unidirectional links of a bi-directional telephony link;

[0011] Figure 2 - illustrates a process flow executed by the destination terminal in association with the process flow of Figure 1;

[0012] Figure 3 - illustrates a means for conveying the speaker's voice profile to the destination terminal so that the speech reproduced at the destination terminal sounds like the speech of the original speaker;

[0013] Figure 4 - illustrates the process executed by the destination terminal when the process of Figure 3 is executed by the source terminal; and

[0014] Figure 5 - illustrates an embodiment of the invention that creates the speaker's voice profile as the speaker communicates speech through a telephony link.

DETAILED DESCRIPTION OF THE INVENTION

[0015] Although a telephony link provides bi-directional communication between the users, the bi-directional link may be described as two separate links. These links are commonly referred to

as the forward and reverse links. Each link has a communication medium that interconnects a source and a destination terminal. The source terminal communicates user speech to the destination terminal. Therefore, the source terminal for the forward link serves as the destination terminal for the reverse link.

[0016] Figure 1 illustrates a process flow 1 executed by the source terminal of either one of two unidirectional links of a bi-directional telephony link and Figure 2 illustrates the associated process flow 5 executed by the destination terminal. As a user's speech is provided to the source terminal, portions of the speech are sampled and converted 2 to text by voice recognition software. Although throughout the disclosure the information about the content of the speech is described as text, it is not necessary that the speech be converted to text, as any symbolic representation of speech can be used. For instance, each word can be represented by a single digital code instead of a combination of codes which each represent a letter. The important aspect of the invention is that a representation of a word can be communicated with substantially less bandwidth than a digital encoding of the sound of the word.

[0017] The textual representation of the speech portion is digitally encoded and communicated 3 to the destination terminal. After each speech portion has been sampled, converted to text, and communicated to the destination terminal, the process 1 determines 4 whether the communication link has terminated. If the link is terminated, the process 1 ends. Otherwise, the process 1 samples the next consecutive portion of the speech and converts 2 this portion to text. The most recently converted text is communicated 3 to the destination terminal and the process 1, again, determines 4 whether the link has terminated. The sequence of converting 2 the next consecutive portion of the received speech to text and then communicating 3 the converted text to the destination unit is repeated until the link is terminated. Process 1 is begun anew when a link is established.

[0018] For the purpose of simplifying the description and process flowcharts, a link termination, as used in this disclosure, may be any form of momentary or permanent discontinuance of the

speaker's speech or of the telephony link. For example, the link termination may be a pause in the speaker's conversation. In these instances, process 1 is renewed when the speaker's next vocal sound is uttered. In an exemplary embodiment, audio processing software can be used to distinguish between speech and background noise to identify pauses. or breaks in speech. Alternatively, the link termination may be the end of the telephony link and process 1 is not renewed until a new telephony link is established.

[0019] While the source terminal executes process 1, the destination terminal executes process 5. Process 5 begins by receiving 6 an individual portion of the communicated text from the source terminal. The received 6 portion of text is converted 7 to speech using the default speech profile of the destination terminal. The speech that is reproduced at the destination terminal has the vocal sound of the person whose voice served as the model for the speech profile, rather than the sound of the speaker. This reproduced speech is conveyed 8 to the listener before process 5 determines 9 whether the link has terminated. If the link has terminated, process 5 ends. Otherwise, the steps of the process 5 are re-executed for the next consecutive portion of text received from the source terminal. The sequence of receiving 6 the next portion of text, converting 7 the received text to speech, and outputting 8 the reproduced speech is repeated until the link terminates. Process 5 is renewed when a link is established.

[0020] Figure 3 illustrates an embodiment of the method that provides a means for conveying the speaker's own natural voice profile to the destination terminal so that the speech reproduced at the destination terminal sounds like the speech of the original speaker. Process 20 begins by determining 21 whether the speaker's voice profile exists at the source terminal. If the profile does not exist, it is created 22 by having the speaker read text provided by the voice recognition software before the link is established. Once the speaker's voice profile is created 22, it may be stored to memory and subsequently accessed without having to be re-created for every instance of an established link. The profile can also be periodically updated based on additional words spoken and stored as the system is used. Next, process 20 determines 23 whether the speaker's voice profile exists at the destination terminal. This determination may be made according to any

known method. As an exemplary method, the source terminal conveys the profile identification to the destination terminal and the latter terminal responds with an indication of whether it has a copy of the profile.

[0021] If the destination terminal has a copy of the speaker's voice profile, process 20 repeatedly executes the same steps executed by process 1, in Figure 1. First, a portion of the speech provided to the source terminal is sampled and converted 24 to text. Next, the text is communicated 25 to the destination terminal. Lastly, a determination 26 is made as to whether the link has terminated. If so, the process 20 ends. Otherwise, the sequence of converting 24 the next consecutive portion of the received speech to text and then communicating 25 the converted text to the destination unit is repeated until the link is terminated. Process 20 is begun again when a link is re-established.

[0022] If the destination terminal does not have a copy of the speaker's voice profile, as determined in step 23, process 20 communicates the voice profile to the destination terminal. Beginning with step 28, process 20 samples a portion of the speech and converts 28 it to text. Thereafter, the converted text and a portion of the speaker's voice profile are communicated 29 to the destination terminal. Next, a determination is made whether the speaker's voice profile has been completely communicated 30 to the destination terminal. If not, a determination 31 is made whether the link has terminated. Process 20 ends when the link is terminated. If the link has not terminated, then the next consecutive portion of the received speech is sampled and converted 28 to text. Both the converted text and the next remaining portion of the voice profile are communicated 29 to the destination terminal and another determination 30 is made whether the speaker's voice profile has been completely communicated to the destination terminal. Process steps 28-31 are repeated until either the link is terminated or the speaker's voice profile has been completely communicated to the destination terminal. If the link terminates before the voice profile is completely communicated, only the un-sent portion of the profile need be communicated when the link is re-established for this same speaker. Once a determination 30 is made that the voice profile has been completely communicated to the destination terminal, the flow of process

20 transfers to step 26 where a determination is made whether the link has terminated. For the remaining existence of the link, the converted text will be communicated to the destination terminal without a portion of the voice profile being sent also.

[0023] Figure 4 illustrates the process 40 executed by the destination terminal when process 20, of Figure 3, is executed by the source terminal. Process 40 begins by determining 41 whether the destination terminal has a copy of the speaker's voice profile. If so, the source terminal is informed 42 of this fact. Thereafter, the steps 43-46 of process 40 are nearly identical to steps 6-9 of process 5, which is illustrated in Figure 2. First, the destination terminal receives 43 the text communicated to it by the source terminal. Then, the text is converted 44 to speech using the speaker's voice profile. The reproduced speech has the vocal characteristics of the original speaker when it is output 45 by the destination device. After each portion of the received text is converted into a reproduction of the original speech, a determination 46 is made whether the link has terminated. If so, process 40 ends. Otherwise, the next portion of text received from the source terminal is similarly received 43, converted 44 to speech using the speaker's voice profile, and output 45 as a reproduction of the original speech in the speaker's voice. Steps 43-45 are repeated until the link terminates.

[0024] If a determination is made in step 41 that the speaker's voice profile does not exist at the destination terminal, then the flow of process 40 transfers to the branch of steps comprising steps 48-53. In step 48, the destination terminal informs the source terminal that it does not have a copy of the speaker's voice profile. Next, the process repeatedly executes a set of steps 49-51 that is similar to the set of steps 6-8 in process 5 of Figure 2. With regard to process 40, however, the destination terminal receives 49 not only the text communicated by the source terminal but also the communicated portion of the speaker's voice profile. The portion of the received voice profile is stored to memory by the destination device and the received text is converted 50 to speech using the default voice profile of the destination terminal. The reproduced speech has the voice characteristics of the person whose voice was used to model the voice profile, rather than the speaker's voice characteristics. As the text is converted to speech, it is

conveyed 51 to the listener. Once the speech portion is output, a determination 52 is made whether the speaker's profile has been completely received from the source terminal. If so, all text subsequently received within the duration of the link will be converted to speech using the speaker's voice profile. Therefore, an affirmative determination in step 52 leads the process out of the branch of steps 49-53 and into the branch of steps 42-46, the destination terminal informs the source terminal that the profile has ben fully received and the speaker's voice profile is used for the speech conversion, steps 43-46.

[0025] If a determination is made in step 52 that the speaker's profile has not been fully received from the source terminal, then a determination 53 is made whether the link has terminated. If so, process 40 ends. Since the received portion of the speaker's voice profile has been stored to memory by the destination terminal, only the remaining portion of the profile need be communicated to the destination terminal when the link is re-established for the same speaker. If the link has not terminated, as determined in step 53, the sequence of steps 49-51 is repeated until either the speaker's voice profile is completely received or the link terminates. This sequence comprises receiving the communicated text and speaker profile portions 49, storing the received portion of the speaker profile to memory, converting the received text to speech 50 using the default voice profile, and outputting the reproduced speech 51.

[0026] Figure 5 illustrates an embodiment of the invention that creates the speaker's voice profile as the speaker communicates speech through the link. Process 60 begins by determining 61 whether the speaker's voice profile exists at the destination terminal. If the voice profile exists at the destination terminal, a portion of the incoming speech provided by the speaker is sampled and converted 62 to text by voice recognition software. The converted text is communicated to the destination terminal before a determination 54 is made whether the link has terminated. If the link has terminated, process 60 ends. If the link has not terminated, the next consecutive portion of the speaker's incoming speech is sampled and converted 62 to text. Again, the converted text is communicated 63 to the destination terminal before a determination 64 is made whether the link has terminated. This sequence of sampling and converting 62 to text the next consecutive portion

of incoming speech, communicating 63 the converted text, and determining 64 whether the link has terminated is repeated until the link terminates.

[0027] If a determination is made that the destination device does not have a copy of the speaker's voice profile in step 61, a determination 66 is made whether the source terminal has a copy of the profile. If so, process 60 executes a branch of steps 67-70, to communicate the speaker's voice profile, that is nearly identical to the branch of steps 28-31 in process 20, as illustrated in Figure 3. Process 60 samples a portion of the speaker's incoming speech and converts 67 it to text. Thereafter, the converted text and a portion of the speaker's voice profile are communicated 68 to the destination terminal. Next, a determination is made whether the speaker's voice profile has been completely communicated 69 to the destination terminal. If not, a determination 70 is made whether the link has terminated. Process 60 ends when the link is terminated. If the link has not terminated, then the next consecutive portion of the speaker's incoming speech is sampled and converted 67 to text. Both the converted text and the next remaining portion of the voice profile are communicated 68 to the destination terminal and another determination 69 is made whether the speaker's voice profile has been completely communicated to the destination terminal. Process 60 steps 67-70 are repeated until either the link is terminated or the speaker's voice profile has been completely communicated to the destination terminal. If the link terminates before the voice profile is completely communicated, only the portions of the profile that were not communicated will be communicated when the link is re-established for this same speaker. Once a determination 69 is made that the voice profile has been completely communicated to the destination terminal, the flow of process 60 transfers to step 64 where a determination is made whether the link has terminated. For the remaining period of the link, the converted text will be communicated to the destination terminal without a portion of the voice profile being sent along with it.

[0028] If the speaker's voice profile does not exist at the destination terminal, as determined by step 66, steps 71-75 are repeatedly executed until either the speaker's profile has been created and completely communicated to the destination terminal or the link is terminated. Beginning with

step 71, a portion of the speaker's speech is sampled and converted 71 to text by voice recognition software. The sampled portion of the speech is also used by the voice recognition software to generate 72 the speaker's voice profile. Both the converted text and any portion of the voice profile available for conveyance are communicated 73 to the destination terminal. It is not necessary to convey portions of the speaker profile as they are created to be within the scope of the invention. It may be advantageous in some implementations to delay transmission until the speaker's profile is developed to a certain extent or until it is fully developed.

[0029] Next, a determination 74 is made whether the speaker's voice profile has been completely generated based upon the voice sampling. If so, no further generation of the speaker's voice profile will be made, for the remaining period of the communication link. The process 60 flow branches to step 69 for a determination of whether the speaker's profile has been fully received at the destination. If the speaker profile has been completely received, the process flows through blocks 67 -70. Once the profile is fully received, process 60 repeatedly samples and converts 62 the next incoming portion of the speaker's incoming speech to text, communicates 63 the converted text to the destination terminal, and re-evaluates 64 whether the link has terminated.

[0030] If a negative determination is made in step 74 as to whether the speaker's voice profile has been completely communicated, a determination 75 is made whether the link has terminated. If so, process 60 is terminated. Any portion of the speaker's voice profile that was not completely generated 72 and communicated 73 to the destination terminal prior to a link termination may be generated and communicated when the next link between these source and destination terminals is established for this speaker. If a negative determination is made in step 75 regarding the termination of the link, the next consecutive portion of the speaker's incoming speech is sampled and converted 71 to text. This next sample is used to continue generating 72 the speaker's voice profile. The available converted text and portion of the generated 72 voice profile are communicated 73 to the destination terminal before another determination is made whether the voice profile has been completely generated. The sequence of sampling the next consecutive portion of the speaker's incoming speech and converting 71 it to text, continuing the generation

72 of the speaker's voice profile, and communicating 73 the converted text and the additional voice profile information is repeated until either the link is terminated or the speaker's voice profile has been completely generated.

[0031] Figure 4 illustrates the process 40 executed by the destination terminal when process 60, of Figure 5, is executed by the source terminal. The same process 40 is executed by the destination terminal when either of processes 20 or 60 are executed by the source terminal.

[0032] The speaker's profile may be periodically updated as use of the system will add words, phrases, proper names, inflections intonations and the like to the database of the user. The updated profile can be used by the source terminal to increase the accuracy of speech recognition. Further, the updated profile can be periodically transmitted to the destination terminal to update a previously transmitted speaker profile to allow for enhanced playback. Updating can take place during a subsequent conversation between the source and destination terminals or can take place as part of a separate transfer connection.

[0033] Because many varying and different embodiments may be made within the scope of the inventive concept herein taught, and because many modifications may be made in the embodiments herein detailed in accordance with the descriptive requirements of the law, it is to be understood that the details herein are to be interpreted as illustrative and not in a limiting sense.